

Enrichissement de l'OCS GE2 à l'aide de l'imagerie satellitaire par A2S

Réunion de démarrage – 06/02/2020 10h-13h

Siège de la Région à Strasbourg (Salle KOEHLIN)

Participants :

CHEIPPE	Xavier	DREAL
WEIBEL	Laurent	SGARE
CARMAUX	Nadia	PNRVN
KAEMPF	Stéphanie	Région (Observatoire régional de la biodiversité)
SCHOTT	Marie-Christine	Région
HAMM	Amandine	Région (visio)
BODAR	Christophe	CA Portes de France Thionville (visio)
SOMMA	Laurent	Pôle métropolitain Nancy Sud Lorraine (visio)
PUISSANT	Anne	A2S
ALLENBACH	Bernard	A2S
DEPREZ	Aline	A2S
MALET	Jean-Philippe	A2S
CAMA	Guillaume	CEN Lorraine
LOPEZ	Joséphine	AFB
VALAT	Florent	DREAL (Observatoire régional du foncier)
HANCZYCK	Jean-Marc	DDT 51
MARTIN	Stéphane	ADEUS
HAIMOUDY	Smaïle	Metz Métropole
GROSSHANS	Yannick	DRAAF

Après un rapide tour de table, l'ordre du jour est rappelé (cf supports de présentation joints):

- Introduction et rappel du contexte
- Présentation de la plateforme A2S, des images satellitaires et des méthodes de traitement
- Présentation des travaux d'enrichissement de l'OCS GE2 : objectifs, méthodes et calendrier

Exemples d'application puis Focus sur chaque poste à enrichir : priorisation, définition du poste, identification des données disponibles pour alimenter les méthodes ou valider les résultats

- Conclusions et perspectives

Introduction

Le contexte des travaux est rappelé (*cf. support de présentation*)

Présentation de la plateforme A2S, des images satellitaires et des méthodes de traitement

Cf. support de présentation : 2020-02-06_Demarrage_A2S-Region_Lot5-FINAL.pdf

Les capacités et compétences de la plateforme sont explicitées. Il s'agit d'exploiter les images satellitaires (principalement) par des calculs de masse basés sur le machine learning et l'IA afin de détecter les comportements et changements. Les moyens sont adossés au datacenter de l'université et permettent de réaliser des calculs sur le périmètre européen en 5 jours en moyenne. La plateforme A2S dispose de 1040 coeurs de calcul dédiés et d'une capacité de stockage de 1,2 pétaoctets (il y a 1024 téraoctets (To) dans un pétaoctet).

Il est essentiel de retenir que les caractéristiques des satellites (et donc des images) conditionnent leur usage selon les besoins.

Généralement les images de bonne résolution (allant dans le panchromatique jusqu'à 1,5 m pour SPOT et 0.70 m pour Pléiades) ont une fréquence annuelle alors que les images de résolution moindre (10m pour Sentinel 1 et 2) peuvent être disponibles jusqu'à tous les 5 jours. La résolution spectrale, allant de 5 bandes pour Pléiades à 13 bandes pour Sentinel 2, ajoute à la capacité discriminante des images. À noter que la qualité des images (selon les conditions météorologiques notamment) est un facteur déterminant de leur bonne exploitation.

Par ailleurs, la haute fréquence est exploitable depuis 2016 alors que les images SPOT et Pléiades sont en libre accès pour les institutionnels adhérents (services publics) depuis 2010. La couverture de la région Grand Est nécessite une mosaïque de 16 images Pléiades (dalles couvrant 100 km x 100 km). Les données de la constellation Sentinel sont libres et accessibles gratuitement.

Les méthodes de traitements utilisées sont des algorithmes de classification fondées sur l'intelligence artificielle. A2S a développé sa propre chaîne de traitement appelée ImCLASS. Des données d'entraînement sont injectées dans un modèle et une série d'attributs particuliers (spectraux, spatiaux, texturaux...) leur est associée. Le modèle sera dès lors en mesure de discriminer la classe recherchée des autres occupations du sol sur l'ensemble de l'image. La classification peut être effectuée soit par pixels soit par segments. Dans ce dernier cas, l'image doit passer par une étape de segmentation (regroupement des pixels de signatures proches).

A titre d'exemple, un attribut spectral pouvant être calculé est l'indice NDV (à partir d'images optiques avec les bandes du rouge et du proche infra-rouge). Cet indice est associé à l'activité chlorophyllienne de la végétation. La valeur de NDVI à une date donnée et son évolution dans le temps sera donc un indicateur pour aider à discriminer des zones végétalisées ou non.

Le résultat de la classification est une carte de probabilité d'appartenance à la classe recherchée. Il est possible de binariser cette carte à partir d'un calcul du seuil de probabilité optimal permettant de séparer au mieux les 2 classes (objet recherché / autre occupation du sol). Différents calculs de seuil peuvent être effectués suivant que l'on cherche à privilégier la précision (on veut être sûr que ce que l'on cartographie appartient effectivement à la classe recherchée quitte à omettre quelques éléments) ou le rappel (on veut être sûr de n'oublier aucun élément de la classe recherchée quitte à faussement cartographier des éléments d'autres classes).

Les livrables peuvent être fournis tant en raster qu'en vecteur.

La chaîne de traitement ImCLASS, validée pour des applications pilotes de l'ESA, permet de fournir en sortie une image sous forme d'une grille de pixels classifiés. La chaîne de traitement a été testée avec succès pour des applications de détection des glissements de terrains et de détection de chantiers en zone urbaine, à des fins de contrôles administratifs.

Il est nécessaire d'alimenter les modèles de classification avec des données d'entrée tels que des relevés terrain, la future OCS GE2 ou encore des inventaires locaux. Mais malgré la mise à disposition de tels échantillons supervisés qui représentent exactement ce qu'on veut extraire, il subsistera en sortie un degré d'incertitude mesurée par une matrice de confusion qui confronte la valeur prédite (issue du processus) à la valeur de l'échantillon (valeur exacte) et une probabilité d'appartenance à la classe recherchée.

Présentation des travaux d'enrichissement de l'OCS GE2 : objectifs, méthodes et calendrier

Il s'agit de confirmer les priorités identifiées sur : la tache artificialisée, les prairies, les vergers, les haies et bosquets.

Tache artificialisée : il est nécessaire de définir ce qu'on entend par tache artificialisée et de définir la fréquence de suivi : annuelle ? bi-annuelle ? trimestrielle sur des secteurs à enjeux ?

Le lien avec les travaux de l'observatoire régional du foncier doit être assuré. À cet effet, le glossaire établi par leurs soins sera transmis à A2S.

L'ADEUS s'interroge sur la plus-value de la démarche par rapport aux données OSO (données vectorielles de Theia, issues de Sentinel 2), exploitées par l'agence d'urbanisme de Metz.

La démarche de production de cette base est différente à plusieurs titres : la temporalité est différente. En effet, les données OSO sont produites annuellement (à partir d'une synthèse d'images pas nécessairement acquises à la même période de l'année) et il est à noter que la nomenclature évolue tous les ans, ce qui entraîne des problèmes de comparabilité. Enfin des confusions entre tache artificialisée et vignes (dans le Piémont...) ont été constatées. Les données sont par ailleurs très pixellisées.

Dans le cadre des présents travaux, il s'agit de produire une tâche identique depuis un T0 (2017), selon une fréquence qui pourrait être infra-annuelle selon les besoins. Les données seront également unifiées pour assurer la possibilité d'une étude de l'évolution temporelle de classes identifiées. La chaîne de production peut travailler en mono-classe et selon des temporalités propres à chaque classe.

Il est également impératif de préserver la permanence de la nomenclature de référence.

Des premiers tests vont être engagés sur certains secteurs : Strasbourg, Nancy, Metz et la Marne (SCOT de Châlons – à confirmer par la DDT 51). Des données test (échantillons d'entraînement) seront transmis au plus vite à la plateforme A2S afin de permettre le traitement et la livraison des premiers résultats (via site ftp), d'ici mi-mai, afin de laisser les partenaires se les approprier et faire un retour pour la réunion du 11/06.

Une différence notable avec OSO (ou toute autre nomenclature) est que l'entrée des travaux se fait par poste. La méthode est adaptée au besoin pour chaque poste et on ne génère pas une nomenclature à partir d'une méthode uniforme pour chaque poste. Il s'agit bien de disposer de bases de données complémentaires spécifiques sur chaque milieu.

La question de la géométrie est également soulignée. Il est nécessaire de disposer des données selon une géométrie harmonisée pour garantir leur bonne exploitation et faciliter leur agrégation. Il est proposé de disposer des données brutes mais également de produits d'usages à différentes échelles et à différents pas de temps pour que les acteurs intéressés puissent en évaluer l'intérêt, suivre les produits sur différents secteurs, par rapport à différentes morphologies urbaines, etc. Le format des données carroyées (INSEE) est particulièrement intéressant à ce titre.

Le SCOT métropole de Nancy, l'ADEUS, la DDT 51 et Metz Métropole sont volontaires pour s'impliquer tout particulièrement sur cette thématique, au sein du groupe projet.

Prairies :

De nombreux enjeux existent autour de l'identification des prairies permanentes et temporaires. De nombreux travaux, peu concluants, existent sur le sujet. Il apparaît difficile à ce stade de pouvoir répondre

à cette identification, d'autant plus que la disponibilité des images ne permet pas d'avoir le recul nécessaire (au moins 6 ans).

Le PNR des Vosges du Nord s'interroge sur la capacité d'A2S à qualifier la qualité des prairies (mono ou pluri espèces). Des données qualifiant tout type de prairie existent au niveau du massif vosgien (travail de terrain dans le cadre d'une thèse CIFRE).

Le CEN Lorraine investigate ces questions et modélise les prairies. Un travail de terrain très important est mené depuis 2011 (1000 parcelles sont déjà prospectées). Il s'agit notamment de distinguer les prairies sèches des prairies humides ou soumises aux crues (comme celles de 2018) et aux variations climatiques.

Le PNRVN s'intéresse à identifier la date de premier changement de fauche de l'année. A2S précise que celle-ci étant sujette à interprétation, elle ne pourra être considérée comme un indicateur en tant que tel.

La dimension agricole devrait pouvoir être caractérisée : il serait intéressant de différencier les pâtures principales et celles qui sont fauchées après le 15 juin. Les pratiques de gestion (trop diverses et dépendantes des conditions climatiques) ne peuvent être identifiées à priori, mais on peut travailler sur la notion de changements intervenus sur une parcelle à un moment donné (les 13 bandes spectrales de Sentinel 2 peuvent y contribuer).

À noter que les trois conservatoires Botaniques des anciennes régions (évoluant à terme vers un Conservatoire Botanique National Nord-Est) possèdent beaucoup de données, notamment sur le massif vosgien, dont nombre sont qualifiées au niveau parcellaire.

Le contexte territorial devrait être pris également en compte. En effet les caractéristiques des prairies diffèrent selon leur localisation et les aléas climatiques. On peut imaginer des modèles différents selon le contexte paysager.

Une proposition est faite pour détecter les prairies qui subissent des crues.

Des tests seront à réaliser en lien avec le PNR des Vosges du Nord et le CEN Lorraine, les emprises restant à définir.

Haies et bosquets :

Besoin notamment au niveau de l'observatoire régional de la biodiversité de disposer d'une information fine et actualisée du linéaire de haies pour le suivi des corridors, dans le cadre de la stratégie régionale. Tests à effectuer sur les images spot 6.

Le PNRVN indique que l'extraction de végétation est un indicateur de l'évolution des milieux ouverts. Spot 6 pourra ainsi apporter à la BD OCS une couche d'information précieuse sur les éléments arborés.

Vergers :

D'après les premiers résultats de réalisation de la BD OCS GE, il semblerait qu'elle réponde de façon intéressante à la problématique de repérage des vergers. A vérifier de façon plus précise par les partenaires concernés par cette thématique.

Le besoin de distinction entre vergers traditionnels et intensifs est également identifié par l'observatoire régional de la biodiversité. A2S indique toutefois la difficulté à traiter ce sujet lorsque les outils à disposition ne comportent pas d'analyse structurelle de la répartition spatiale des arbres constituant ce type de plantation.

L'existence d'une étude sur les vergers du Sundgau est néanmoins évoquée.

Conclusion

A2S va engager de premières explorations sur les postes identifiés (et en premier lieu sur la tache artificialisée), à commencer par collecter les données devant alimenter le modèle. Les résultats seront transmis au groupe projet d'ici la mi-mai, la prochaine réunion étant fixée au jeudi 11 juin de 10h à 13h à Metz (Hôtel de Région).

Les supports de présentation seront transmis aux participants avec le compte-rendu. Ils seront également disponibles sur le cloud dédié au projet (<http://www.osge2.eu/accueil.html>) et sur la page projet GéoGrandEst (<https://www.geograndest.fr/portail/fr/projets/occupation-du-sol>).

Pour les partenaires désirant s'impliquer sur l'un des postes évoqués, un document sera partagé afin de :

- ✓ s'accorder sur les termes => vers un glossaire partagé (à valider lors de la prochaine réunion)
- ✓ disposer des informations pour démarrer les tests, décrivant : (1) les secteurs d'étude (avec shapefile associé), (2) les données de référence / validation => **deadline semaine du 09/03**
- ✓ identifier, pour chaque objet d'étude, le(s) usage(s) attendu(s) avec des spécifications sur l'échelle de représentation, la surface minimum de l'objet, .. à compléter

Pour accéder à ce document (pour les partenaires non-représentés lors de cette réunion de démarrage), merci d'envoyer un mail à : anne.puissant@unistra.fr